Residual kriging for functional data. Application to the spatial prediction of salinity curves

Adriana Reyes[1], Ramón Giraldo[1,*]and Jorge Mateu[3]

[1]Universidad Nacional de Colombia, Bogotá, Colombia.

[2]Universitat Jaume I, Castellón, Spain.

Abstract

Recently several methodologies for carrying out geostatistical analysis of functional data has been proposed. All of them assume that the spatial functional process considered is stationary. However in practice we often have nonstationary functional data sets because there is spatial trend in the mean. Here we propose a methodology to extend kriging predictors for functional data to the case where the mean function is not constant through the region of interest. We consider an approach based on the classical residual kriging method used in univariate geostatistcs. We propose a three steps procedure. Initially a functional regression model is used for detrending the mean. Posteriorly we apply kriging methods for functional data to the regression residuals for doing prediction of a residual curve on a non-data location. Finally the prediction curve is obtained as the sum of the trend and the residual prediction. We apply the methodology to a salinity data set corresponding to 21 salinity curves recorded a the Ciénaga Grande de Santa Marta estuary, located in the Caribbean coast of Colombia. A cross-validation analysis was carried out in order to establish the performance of the methodology proposed.

Keywords: Cross-validation; Functional linear model; Residual kriging; Salinity.

1 Introduction

In the last years the number of situations where the data to be analyzed are functions have increased. Since beginning of the nineties, functional data analysis (FDA) (Ramsay and Silverman, 2005) has been used to describe, analyze and model this kind of data. Functional versions for a wide range of statistical tools (ranging from exploratory and descriptive data analysis to linear models and multivariate techniques) have been developed (see an overview in González-Manteiga and Vieu

*Corresponding author. E-mail: rgiraldoh@unal.edu.co. Fax:+57-1-6437361

(2007)). The standard statistical techniques for FDA are focused on assuming independence among functions. However, in several disciplines of applied sciences there exists interest in modeling correlated functional data: this is the case when functions are observed over a discrete set of time points or when these functions are observed in different sites of a region. In this paper we focus on spatially correlated functional data, and particularly in predicting curves in sites of a region with spatial continuity. Several works have been devoted to solve this problem. Giraldo et al. (2011) propose an ordinary kriging predictor for functional data whose parameters, as in the univariate case, are scalars. Other approaches are based on consider kriging predictors with functional parameters (Giraldo et al., 2010; Nerini et al., 2010). All of these methods assume that the spatial functional process considered is stationary, that is, the mean function is constant (no trend), the variance functions is constant, and the covariance function depends on the distance between the locations. However in many practical applications, particularly when we analyze environmental data, the assumption of constant mean function could be not realistic.

In classical geostatistcs there are several alternatives to solve the problem of spatial prediction for nonstationary process because the mean is not constant. These are very similar. In others, universal kriging (UK), a variant of UK called in the geostatistical literature kriging with external drift (KED), and residual kriging(RK) are used for doing spatial prediction when there is a trend in the mean. All of them are based on the estimation of regression models with spatially correlated errors. The semivariance and covariance must be estimated from the detrended data (Gotway and Hartford, 1996). UK is the name used when the coordinates are the predictors. If some auxiliary variables are used, rather than the coordinates, the term KED is preferred. In the case of UK or KED, the predictors are included in the kriging solution system and additional unbiasedness constraints must be are made. If the drift and regression residuals are considered separately, and ordinary kriging based on residuals and the trend are summed the method is called RK (this is also known as regression-kriging). All of these methods have theoretical problems when ordinary least squares (OLS) is used to estimate the regression parameters because the estimation of the semivariance based on residuals is biased. To solve this problem Cressie (1993) propose to use median-polish kriging. This procedure is similar to RK but in this case the trend is estimated by median-polish and ordinary kriging is applied on the median-polish residuals. Also restricted maximum likelihood method (REML) can be used to estimate from the data both the trend model and the covariance parameters of the residual process (Gotway and Hartford, 1996; Minasny and McBratney, 2007). Another possibility is is to compute the semivariogram from studentized or recursive OLS residuals (Schabenberger and Gotway, 2005). All of these approaches allow to solve the problem of using OLS residuals to estimate the semivariance. However some based on real data studies has reveal that the difference between UK, KED, and RK based on OLS residuals with other methods is not large (Knotters et al., 1995; Minasny and McBratney, 2007). This is consequence that the bias of a residuals-based variogram estimator is small at lags near to the origen but more substantial at distant lags and because the kriging is carried out in local neighborhoods, the fitted variogram is evaluated at smaller lags, precisely where it has been well fitted (Cressie, 1993). From a practical point of view RK is easier than UK or KED and has proven to be a robust technique for practical applications (Minasny and McBratney, 2007). This methodology has been widely applied in modeling environmental data (see, for example Knotters et al. (1995), and Alsamamra et al. (2009)). In this work we consider the extension of RK to the case where data are functions, that is, when we have a spatially correlated data set where the mean function is not constant through the region of interest. Initially a functional regression model is used for detrending the mean. OLS method is used to estimate the functional parameters (Ramsay and Silverman, 2005). Posteriorly we apply kriging methods for functional data (Delicado et al., 2010) to the OLS regression residuals for doing prediction of a residual curve on a non-data location. Finally the prediction curve is obtained as the sum of the trend and the residual prediction. This approach do not consider bias correction because the use of maximum likelihood (ML) methods for estimating parameters of functional regression models is to the best of our knowledge an open problem in FDA, particularly when we have a functional regression model with functional response. This topic is posteriorly treated in Section 4. We apply the approach to a real data set corresponding to salinity curves obtained at 21 monitoring stations of the Ciénaga Grande de Santa estuary located in the Caribbean coast of Colombia. A cross-validation analysis was carried out. The results show that the methodology proposed is a good alternative for doing spatial prediction of functional data when the mean functions is not constant through the region of interest. A comparison between three alternatives of doing kriging prediction with OLS regression residuals were considered. Results indicate that the most simple method (based on ordinary kriging for functional data) is the best option in this case.

This work is organized as follows. Section 2 gives a brief overview about kriging predictors for functional data and describes the methodology proposed. In Section 3 an application with real data is shown. The paper ends with a brief discussion and suggestions for further research.

2 Residual kriging for functional data

In this section we show the basics of three methods for carrying out kriging prediction of functional data assuming stationarity. Then we show how these ones can be extended to the nonstationary case defining the residual kriging predictor for functional data.

Let $\left\{\boldsymbol{\chi}_s(t), t \in T, s \in D \subset \mathbb{R}^d\right\}$ be a *spatial functional process* (Delicado et al., 2010) defined on some compact set T of $\mathbb{R}$ where s is a generic data location in the d-dimensional Euclidean space (d is usually equal to 2) and $\boldsymbol{\chi}_s(t)$ are *functional random variables* (Ferraty and Vieu, 2006), defined as random elements taking values in an infinite dimensional space (or functional space). We assume that this one is a separable Hilbert space $\boldsymbol{H}$ of square integrable functions defined on T. The kriging predictors for functional data (Delicado et al., 2010) assume that the functional random process is second-order stationary and isotropic, that is, the mean and variance functions are constant and the covariance and semivariance depends only on the distance between sampling points. Formally is assumed that

- $E(\boldsymbol{\chi}_s(t)) = \mu(t)$
- $V(\boldsymbol{\chi}_s(t)) = \sigma^2(t)$
- $\mathrm{Cov}(\boldsymbol{\chi}_{s_i}(t), \boldsymbol{\chi}_{s_j}(u)) = C(h;t,u),\ s_i, s_j \in D, t, u \in T,\ h = \|s_i - s_j\|$, the Euclidean distance. If $t = u$, $\mathrm{Cov}(\boldsymbol{\chi}_{s_i}(t), \boldsymbol{\chi}_{s_j}(t)) = C(h;t)$.
- $\frac{1}{2}\mathrm{V}(\boldsymbol{\chi}_{s_i}(t) - \boldsymbol{\chi}_{s_j}(u)) = \gamma(h;t,u),\ s_i, s_j \in D, t, u \in T,\ h = \|s_i - s_j\|$. If $t = u$, $\frac{1}{2}\mathrm{V}(\boldsymbol{\chi}_{s_i}(t) - \boldsymbol{\chi}_{s_j}(t)) = \gamma(h;t)$.

Three kriging predictors based on stationarity are considered here. The simplest called *ordinary for functional data* (Giraldo et al., 2011) has the same expression of a classical kriging predictor but considering curves instead of data. This predictor is defined as

$$\hat{\boldsymbol{\chi}}_{s_0}(t) = \sum_{i=1}^{n} \lambda_i \boldsymbol{\chi}_{s_i}(t),\ \ \lambda_1, \ldots, \lambda_n \in \mathbb{R}. \tag{1}$$

The parameters $\lambda_i,\ i = 1, \ldots, n$ are found as the solution of the linear system

$$\begin{pmatrix} \int_T \gamma_{s_1 s_1}(t)dt & \cdots & \int_T \gamma_{s_1 s_n}(t)dt & 1 \\ \vdots & \ddots & \vdots & \vdots \\ \int_T \gamma_{s_n s_1}(t)dt & \cdots & \int_T \gamma_{s_n s_n}(t)dt & 1 \\ 1 & \cdots & 1 & 0 \end{pmatrix} \begin{pmatrix} \lambda_1 \\ \vdots \\ \lambda_n \\ -\mu \end{pmatrix} = \begin{pmatrix} \int_T \gamma_{s_0 s_1}(t)dt \\ \vdots \\ \int_T \gamma_{s_0 s_n}(t)dt \\ 1 \end{pmatrix},$$

where the function $\gamma(h) = \int_T \gamma_{s_i s_j}(t)dt$, $h = \|s_i - s_j\|$, is called trace-semivariogram. This function is estimated by

$$\hat{\gamma}(h) = \frac{1}{2|N(h)|} \sum_{i,j \in N(h)} \int_T (\chi_{s_i}(t) - \chi_{s_j}(t))^2 dt, \tag{2}$$

where $N(h) = \{(s_i, s_j) : \|s_i - s_j\| = h\}$, and $|N(h)|$ is the number of distinct elements in $N(h)$. Once we have estimated the trace-semivariogram for a sequence of K values h_k, we propose to fit a parametric model (any of the classical and widely used models such as spherical, Gaussian, exponential or Matérn could well be used) to the points $(h_k, \hat{\gamma}(h_k)),\ k = 1, \ldots, K$, as if they were obtained in the classic geostatistical setting.

A second alternative is to consider a kriging predictor with functional parameters where the influence of curves on the prediction is given in the same argument $t \in T$. This predictor is considered in Giraldo et al. (2010) and is called *continuous time-varying kriging for functional data.* This predictor is defined by the expression

$$\boldsymbol{\chi}_{s_0}(t) = \sum_{i=1}^{n} \lambda_i(t) \boldsymbol{\chi}_{s_i}(t), t \in T,\ \lambda_i : T \mapsto \mathbb{R},\ i = 1, \ldots, n. \tag{3}$$

The estimation of the functional parameters $\lambda_i(t), i = 1, \ldots, n$, is carried out by using an approach based on the use of basis functions. The curves $\boldsymbol{\chi}_{s_i}(t)$ and the functional parameters are represented in terms of K basis functions. Thus a new parametrization is given where we have K new variables (formed with the coefficients of fitting the basis functions to the data) and $n \times K$ parameters (given by coefficients of fitting the basis function to the functional parameters). In order to estimate the parameters a linear model of coregionalization between the new variables must be estimated. A detailed overview to this approach is given in Giraldo et al. (2010).

A third alternative for functional kriging is allowing the functional parameters to be defined in $T \times T$. Then, the predictor of $\boldsymbol{\chi}_{s_0}(t)$ is

$$\hat{\boldsymbol{\chi}}_{s_0}(v) = \sum_{i=1}^{n} \int_T \lambda_i(t,v)\boldsymbol{\chi}_{s_i}(t)dt, \; v \in T, \tag{4}$$

such that $\lambda_1(t,v), \ldots, \lambda_n(t,v) : T \times T \to \mathbb{R}$. This kriging predictor has been separately proposed by Giraldo (2009) (called as *functional kriging total model*) and by Nerini et al. (2010) (called as *Cokriging for spatial functional data*). Again the observed functions $\boldsymbol{\chi}_{s_i}(t)$ and the functional parameters $\lambda_i(t,v) i = 1, \ldots, n$ are expanded in terms of K basis functions in order to give a solution to the problem of estimating the functional parameters. An overview about the estimation of parameters in the predictor 4 can be done in Giraldo (2009).

In this work we propose a new predictor which is useful when the mean function is not constant. We give a methodology for doing spatial prediction when the mean function depends on the location, that is, when $E(\boldsymbol{\chi}_s(t)) = \mu_s(t)$. The methodology has three steps. Initially a functional regression model (FRM) (Ramsay and Silverman, 2005) is used for detrending the mean. Posteriorly we apply some kriging method for functional data to the regression residuals for doing prediction of a residual curve on a non-data location. Finally the predicted curve is obtained as the sum of the trend and the residual prediction. Now we show a detailed description of each step.

Ramsay and Silverman (2005) have shown several alternatives for doing functional regression analysis, that is, to estimate functional FRM where either response or predictor variables are functions. The first step of the proposed methodology is to estimate a FRM with functional response and scalar covariates (Ramsay and Silverman, 2005, chapter 13). Specifically we use the model

$$Z_{s_i}(t) = \alpha(t) + \beta_1(t)x_i + \beta_2(t)y_i + \epsilon_i(t), \tag{5}$$

where $Z_i(t), i = 1, \cdots, n$ are the functions at visited locations, (x_i, y_i) are geographical coordinates, $\alpha(t), \beta_1(t), \beta_2(t)$ are the functional parameters of interest and $\epsilon(t)$ is a white noise for each $t \in T$. The parameters are estimated by least squares using an approach based on the use of basis functions (Ramsay and Silverman, 2005). The model in equation (5) has scalar covariates, however other FRM could also be applied (for instance taking functional covariates). Once estimated the regression model in a second step we obtain the residuals

$$\begin{aligned} e_{s_i}(t) &= Z_{s_i}(t) - \hat{Z}_{s_i}(t) \\ &= Z_{s_i}(t) - \left(\hat{\alpha}(t) + \hat{\beta}_1(t)x_i + \hat{\beta}_2(t)y_i\right), \end{aligned} \tag{6}$$

and based on these ones we can do spatial prediction of a residual curve at non-visited locations by using either the predictors defined in equations (1),(3), and (4), respectively, that is, we predict a residual function by using some of these predictors

$$\hat{e}_{s_0}(t) = \sum_{i=1}^{n} \lambda_i e_{s_i}(t), \tag{7}$$

$$\hat{e}_{s_0}(t) = \sum_{i=1}^{n} \lambda_i(t) e_{s_i}(t), \tag{8}$$

or

$$\hat{e}_{s_0}(v) = \sum_{i=1}^{n} \int_T \lambda_i(t,v) e_{s_i}(t) dt, \tag{9}$$

where $e_i(t), i = 1, \ldots, n$ are the residual curves obtained from the estimated FRM.

Finally in a third step the prediction on a non-visited location is achieved by

$$\hat{\boldsymbol{\chi}}_{s_0}(t) = \hat{Z}_0(t) + \hat{e}_0(t), \tag{10}$$

where $\hat{\boldsymbol{\chi}}_{s_0}(t)$ is the function predicted on the location s_0, $\hat{Z}_{s_0}(t) = \hat{\alpha}(t) + \hat{\beta}_1(t)x_0 + \hat{\beta}_2(t)y_0$ is the trend estimated on the location with coordinates (x_0, y_0) and $\hat{e}_0(t)$ is the prediction of a residual function on a non-visited location s_0. We call $\hat{\boldsymbol{\chi}}_{s_0}(t)$ in equation (10) *residual kriging predictor* for functional data.

3 Spatial prediction of salinity curves

In this section we apply the methodology described in Section 2 to a real data set corresponding to salinity data measured at 21 monitoring stations of the lagoonal-estuarine system comprised by the Ciénaga Grande de Santa Marta (CGSM) and Complex of Pajarales (CP) (Figure 1). The CGSM-CP is the largest coastal lagoon system of Colombia. Several works have shown that the salinity is one of the variables that better describes the changes of this ecosystem (Blanco et al., 2006). In semi-closed ecosystems as the CGSM and CP the salinity is highly variable due to in others river discharges, winds, and the movement of tides. Identifying spatial and temporal variability in salinity provide important ecological information and for this reason is important to know its spatial distribution. Classical univariate and multivariate geostatistical methods have been used to get this objective. Here we give a new tool for modeling these data. Our approach allows to predict the spatio-temporal salinity behavior.

Biweekly data from October 1988 to March 1991 were recorded (left panel, Figure 2). Data were smoothed by using a B-splines basis with $K = 15$ functions (Figure 2). The number of basis function was chosen by cross-validation (Ramsay and Silverman, 2005).

According to plots in Figure 2 the salinity level in the monitoring stations of CP (west side in Figure 2) is higher than the salinity level in the CGSM. We note in the plots that two stations at CGSM have considerably lower salinity values than others. These correspond to stations RFU and RSE which are located at the mouths of Fundación and Sevilla rivers, respectively. This result indicates empirically that there is a decreasing trend from west to east. Thus the stationarity assumption with this data set could be not valid and the application of the predictors in equations (1), (3), or (4) for doing spatial prediction of salinity curves on non-visited locations from CGSM and CP could be inappropriate. In this case the application of the residual kriging for functional data seems to be a better option. Taking into account this result we carry out a functional regression analysis with functional response (the smoothed salinity curves) and scalar covariates (the geographical coordinates). We estimate the functional parameters by ordinary least squares following Ramsay and Silverman (2005). We use the R library `fda` (Ramsay et al., 2010) to fulfill this task. Specifically we estimate the model

$$\begin{aligned} Z_i(t) &= \hat{\alpha}(t) + \hat{\beta}_1(t)\text{Longitude}_i + \hat{\beta}_2(t)\text{Latitude}_i + e_i(t) \\ &= \hat{Z}_i(t) + e_i(t) \end{aligned} \tag{11}$$

where $Z_i(t)$ are the smoothed salinity curves (Figure 2, right panel), $\hat{Z}_i(t)$ are the regression estimations of the salinity curves, $\hat{\alpha}(t)$, $\hat{\beta}_1(t)$, and $\hat{\beta}_2(t)$ are the estimations of the functional parameters, and $e_i(t)$ are the residual curves.

A plot with the estimated functional parameters is shown in Figure 3. We observe that parameters are significantly different from zero (only β_1 is significantly equal to zero in some time periods). The estimation of the mean $\hat{\alpha}(t)$ show that the level of salinity increased in the period of study. According to the sign of $\hat{\beta}_1$ we can conclude that the salinity decrease from west to east, that is, when the longitude increase (right panel Figure 1) the salinity level decrease (in those time periods where $\hat{\beta}_1$ has negative sign and is significantly different from zero). This is coherent with the result before mentioned about that the level of salinity is higher in stations from CP than in stations of CGSM (Figure 2). The sign $\hat{\beta}_2$ indicates that the salinity level increase from south to north (Figure 1). This trend is particularly due to the stations located in the south of the system (RFU and BRF, Figure 1) have a low level of salinity because of they have direct influence of the Fundación river.

Once estimated the functional regression model we use the residual regression curves $e_i(t), i = 1, \cdots, 21$ (left panel Figure 4) to do spatial prediction of a residual curve on a non-visited location (with coordinates (945000, 1692000)) by means of the predictors given in equations (7), (8), and (9). As in the case of the salinity curves a functional cross-validation analysis (Ramsay and Silverman, 2005) indicated that a B-splines basis with $K = 15$ functions was appropriated for smoothing the residual data (left panel, Figure 4).

For comparison proposes we also obtained predictions applying directly the predictors in equations (1), (3), and (4) to the salinity data. The predictions obtained by the six methods are shown in right panel of Figure 4. We note that there is some important differences between the predictions. In others, two questions arise naturally after observing this plot. Which method is better?. Methods based on residual kriging are better than methods for stationary data?. In order to solve these ones and to verify the goodness-of-fit of the proposed predictors we use a *functional cross-validation analysis*. Each individual smoothed curve $\chi_{s_i}(t)$, $i = 1, \cdots, 35$, was temporarily removed, and further predicted from the remaining ones. This procedure was realized with each one of the six predictors (three based on the stationarity assumption and three considering the residual kriging approach). We evaluate on $j = 1, \cdots, 55$ the cross-validation predictions obtained by each method. Summary statistics of the sum of squared errors of *functional cross-validation* (SSE_F) resulting of applying the six methods to the data set considered are shown in Table 1.

Table 1: Summary statistics of sum of squared errors of cross-validation. OKFD: Ordinary kriging for function-valued spatial data; CTVFD: Continuous time-varying kriging for functional data; FKTM: Functional kriging; ROKFD: Residual kriging based on OKFD; RCTVFD: Residual kriging based on CTVFD; RFKTM: Residual kriging based on FKTM.

Statistic	OKFD	CTKFD	FKTM	ROKFD	RCTKFD	RFKTM
Minimum	242	316	237	538	501	778
Median	806	1336	723	1777	1715	1770
Mean	2710	2870	2652	2146	2157	2175
Maximum	16840	12063	8483	7313	7215	8483
Standard deviation	4018	3527	3824	1463	1442	1693
Sum	56909	60278	55695	45072	45287	45678

The summary statistics of SSE_F values (Table 1) and in particular the sum of SSE_F values indicate that methods based on residual kriging (ROKFD, RCTVFD, and RFKTM) have better performance than other predictors. We also note that though the differences between methods based on residual kriging are small we get in general (using as indicator the sum of SSE_F values) better results with ROKFD. In all cases LBA located in the CGSM (Figure 1), is the station with the worst prediction. According to the maximum SSE_F values in this station RCTVFD and ROKFD methods have the best behavior. The salinity level in station LBA is as high as the stations located in CP (separated from LBA more than 15 km). The salinity at stations closer to LBA (as RJA and BRS) have a very different behavior across the time. Taking into account in kriging prediction curves from locations closer to the prediction site will have greater influence than other more separated is expected to have bad prediction at this station. The use of the residual kriging prediction, in particular of RCTVFD and ROKFD methods, allows to improve the prediction in this case. We note that there is significantly difference between the maximum of SSE_F values when we compare the methods based on stationarity and the methods based on residual kriging. In the last case, the effect of the estimated functional parameter $\hat{\beta}_2(t)$ in equation 11 (which is positive and indicate that there is a trend from south to north) allows to improve the prediction by kriging. The differences of SSE_F values between ROKFD, RCTVFD and RFKTM are very small in a high proportion of sites. However taking into account that, from a practical and computational point of view, ROKFD is simpler than RCTVFD and RFKTM we consider that in this case ROKFD is the best option.

A plot of cross-validations residuals obtained by ROKFD is shown in Figure 5. The residual standard deviation is greater at the end of the time perido considered (Figure 5) where the smoothed and the predicted curves have greater variation. The residual mean varies around zero which indicates that the predictions obtained by ROKFD are unbiased (Figure 5). Similar results are obtained with RCTVFD and RFKTM.

4 Discussion and further research

We have shown how three predictors used in geostatistcs for functional data, which are based on the stationarity assumption for the mean function, can be generalizad to the nonstationary context (where the mean change inside the study) by using a functional regression model for detrending the mean function. The predictor proposed combine functional regression estimation with functional kriging predictions based on the regression residual in order to do prediction on non-visited locations. We apply the methodology proposed to a data set corresponding to salinity curves obtained in a estuary located in the colombian Caribbean. Results with this data indicates that although several alternatives for doing functional kriging prediction with the regression residuals could be used, the method based on the use of ordinary kriging for functional data is the best alternative. In univariate geostatistics several works have shown that the use of residual kriging (and other alternatives such as UK and KED) with OLS estimation causes biased estimations from the covariance parameters. The solution proposed in that case is based on the use of maximum likelihood estimation for estimating simultaneously both the regression parameters and the covariance parameters. In our approach for doing residual kriging for functional data we propose to estimate the parameters from the functional regression model by OLS. The effect of using OLS in this case must be studied. In particular we need to establish one the one hand, if we use OKFD, the effect of estimating the trace-variogram function by using equation (2) based on functional regression residuals or on the other hand, if we use CTKFD and FKTM, which is the effect of estimating a linear model of coregionalization based on coefficients of fitting a basis function to functional regression residuals. We consider that a simulation study must be done to solve these questions. The alternative of doing maximum likelihood estimation of functional parameters when we have a functional regression model with functional response and spatially correlated errors is an open problem in FDA. Results of cross-validation analysis carried out with the salinity data set show a good performance of the proposed predictor, and indicate from a descriptive point of view that it can be adopted as a valid method for modeling spatially correlated functional data when the mean function is not constant through the region of interest.

References

Alsamamra, H., J. Ruiz-Arias, D. Pozo-Vázquez, and R. Tovar-Pescador (2009). A comparative study of ordinary and residual kriging techniques for mapping global solar radiation over southern spain. *Agricultural and Forest Meteorology 149*, 1343–1357.

Blanco, J., E. Viloria, and J. Narváez (2006). ENSO and salinity changes in the Ciénaga Grande de Santa Marta coastal lagoon system, Colombian Caribbean. *Estuarine, Coastal and Shelf Science 66*, 156–167.

Cressie, N. (1993). *Statistic for Spatial Data.* New York: John Wiley & Sons.

Delicado, P., R. Giraldo, C. Comas, and J. Mateu (2010). Statistics for spatial functional data: some recent contributions. *Environmetrics 21*, 224–239.

Ferraty, F. and P. Vieu (2006). *Non Parametric Functional Data Analysis. Theory and Practice.* New York: Springer.

Giraldo, R. (2009). *Geostatistical analysis of functional data.* Ph. D. thesis, Universitat Politècnica de Catalunya.

Giraldo, R., P. Delicado, and J. Mateu (2010). Continuous time-varying kriging for spatial prediction of functional data: An environmental application. *Journal of Agricultural, Biological, and Environmental Statistics 15*(1), 66–82.

Giraldo, R., P. Delicado, and J. Mateu (2011). Ordinary kriging for function-valued spatial data. *Environmental and Ecological Statistics 18*, 411–426.

González-Manteiga, W. and P. Vieu (2007). Statistics for functional data. *Computational Statistics and Data Analysis 51*, 4788–4792.

Gotway, C. and A. Hartford (1996). Geostatistical methods for incorporating auxiliary information in the prediction of spatial variables. *Journal of Agricultural, Biological and Environmental Statistics 1*(1), 17–39.

Knotters, M., D. Brus, and J. Oude Voshaar (1995). A comparison of kriging, cokriging and kriging combined with regression for spatial interpolation of horizon depth with censored observations. *Geoderma 67*, 227–246.

Minasny, B. and A. McBratney (2007). Spatial prediction of soil properties using EBLUP with the Matérn covariance function. *Mathematical Geology 140*, 324–336.

Nerini, D., P. Monestiez, and C. Manté (2010). Cokriging for spatial functional data. *Journal of Multivariate Analysis 101*, 409–418.

Ramsay, J. and B. Silverman (2005). *Functional data analysis. Second edition.* New York: Springer.

Ramsay, J., H. Wickham, S. Graves, and G. Hooker (2010). ***fda**: Functional Data Analysis. R package version 2.2.6.*

Schabenberger, O. and C. Gotway (2005). *Statistical Methods for Spatial Data Analysis.* Boca Ratón: Chapaman & Hall.

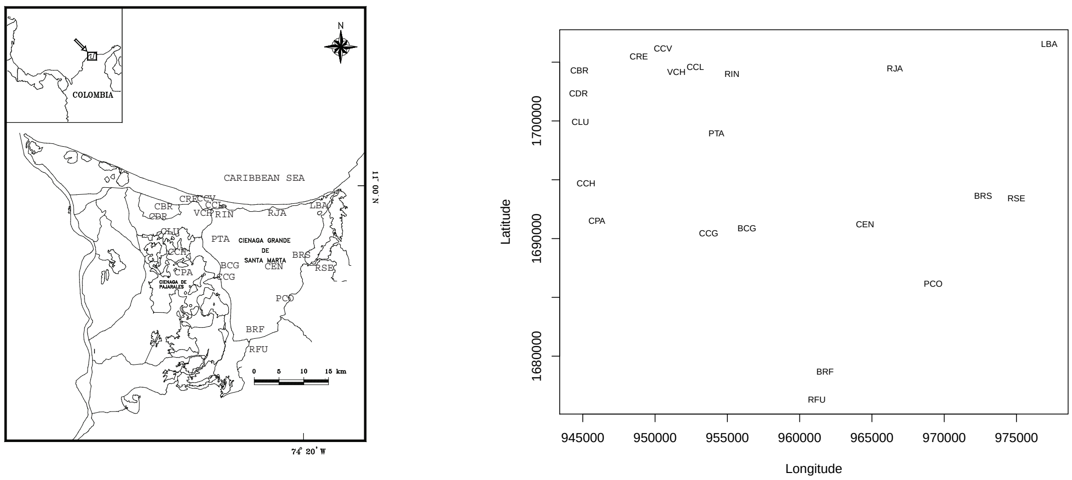

Figure 1: 21 monitoring stations of the lagoonal-estuarine system comprised by the Ciénaga Grande de Santa Marta (stations LBA, BRS, PCO, BRF, RSE, RFU, CEN, RJA, BCG, PTA, RIN) and Complex of Pajarales (stations CCL, VCH, CCV, CRE, CBR, CDR, CLU, CCH, CPA, CCG).

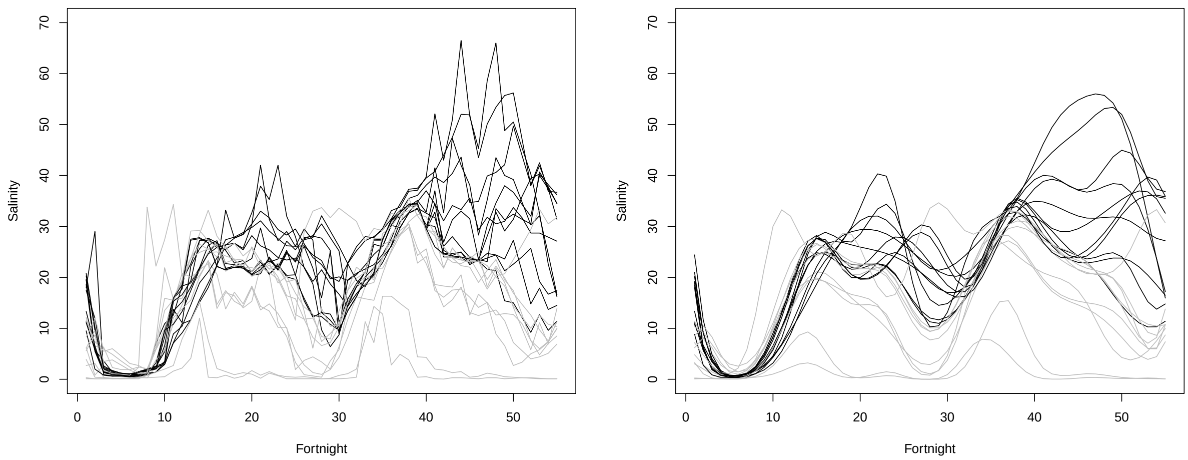

Figure 2: Left panel: Biweekly salinity data recorded on 21 monitoring stations of the lagoonal-estuarine system comprised by Ciénaga Grande de Santa Marta and Complex of Pajarales. Right panel: Salinity data smoothed by using a B-splines basis with 15 functions. In both panels *gray curves* correspond to data measured in Ciénaga Grande de Santa Marta and *dark curves* to data measured in stations from Complex of Pajarales.

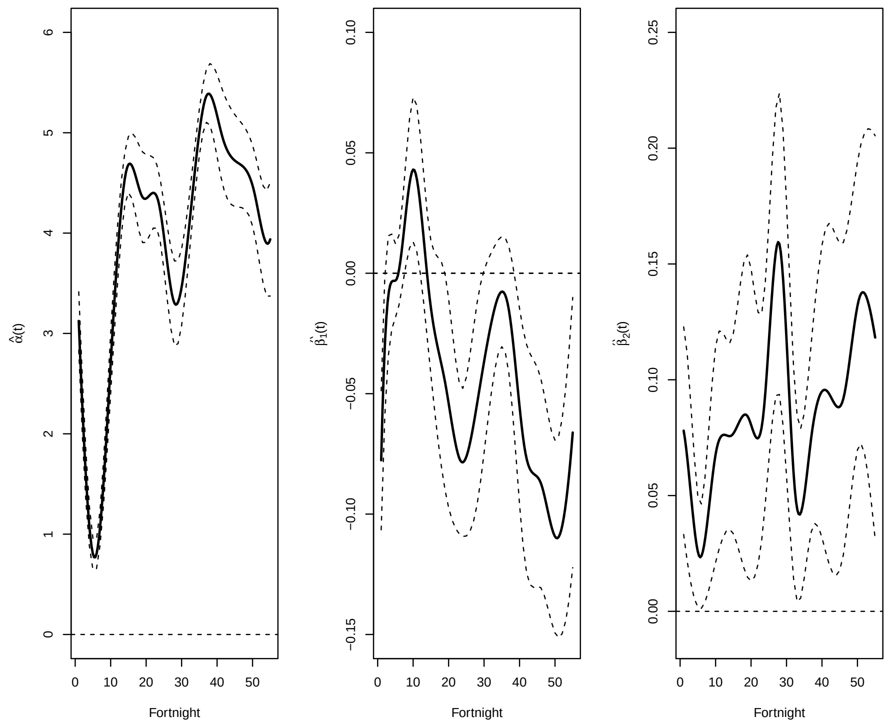

Figure 3: Estimated parameters from the functional regression model between the salinity curves and the geographical coordinates of stations where data were recorded. *Dark lines* correspond to estimations and *dashed lines* to the confidence intervals.

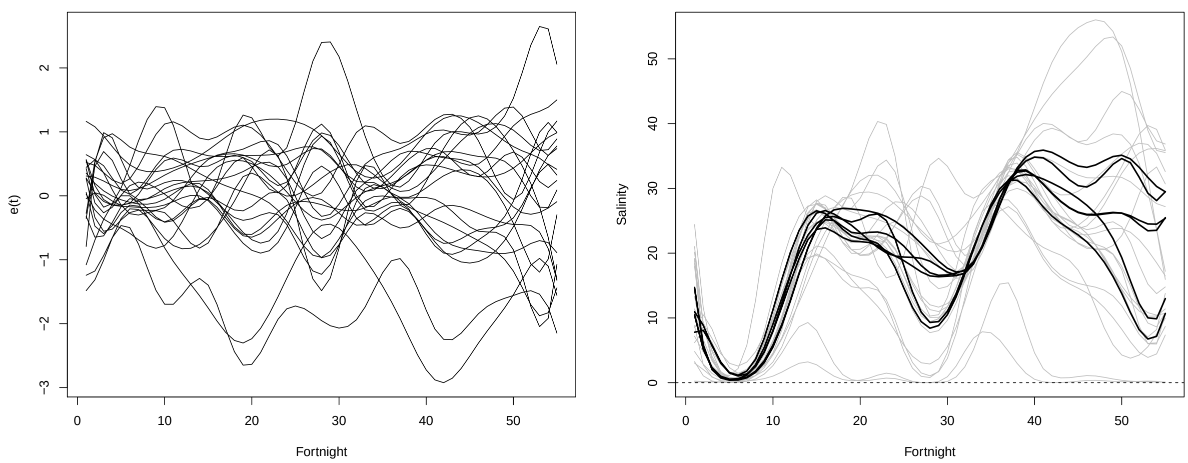

Figure 4: Left panel: Residuals curves of the functional regression model between salinity curves and the geographical coordinates of monitoring stations of Ciénaga Grande de Santa Marta and Complex of Pajarales. Residuals were smoothed by a B-splines basis with 15 functions. Right panel: Predictions of salinity curves on a non-visited location by means of six spatial predictors for functional data (gray lines are smoothed salinity curves and dark lines are the predictions).

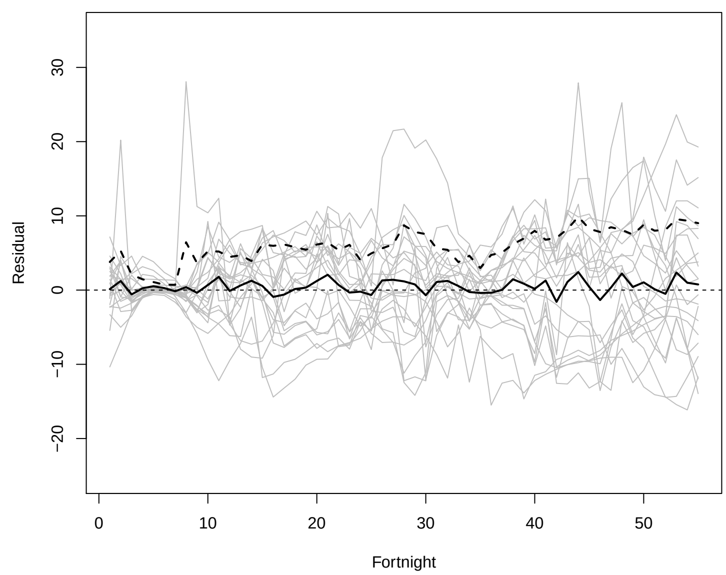

Figure 5: Cross-validation residuals obtained by kriging residual (ROKFD) (gray curves). Dark line is the mean of cross-validation residuals and dashed line is the standard deviation of cross-validation residuals.